

DC14 Emmanuel Sylvain FOMAT

IRP14: Language awareness and communicative ability through exposure: Data-driven approaches with corpora and GenAI

Research fields: Applied Linguistics, Contrastive Linguistics, Corpus Linguistics, Didactics

Co-supervisors of IRP14:

Eva Vetter, eva.vetter@univie.ac.at

Alex Boulton, alex.boulton@univ-lorraine.fr

This PhD position is part of a double-degree programme. Upon completion, candidates will be awarded **one joint doctoral degree**, on two PhD diplomas, following research stays at both institutions: **University of Vienna** (diploma conferred by the **University of Vienna**) and **Université de Lorraine** (diploma conferred by the **Université de Lorraine**).

Scientific context of the Individual Research Project IRP14:

IRP14 is embedded into **Work Package 4 (WP4) ‘Language learning and plurilingualism in the era of GenAI’**. In light of technological advances in digital tools and resources, as well as the ongoing migration crisis, WP4 reconsiders the main concepts and approaches to language learning and teaching: it investigates practical language awareness, metalinguistic awareness and critical language awareness as well as learner autonomy, motivation and plurilingualism from the learners’ perspective. It will conduct pedagogical experiments and analyse interview, survey and classroom data, as well as curricula and language teaching/learning materials. WP4 will produce guidelines for students, teachers, teacher trainers and decision-makers and innovative teaching materials, resources and pathways for formal and non-formal language learning in a digital Europe.

In this context, IRP14 investigates how massive exposure to structured language can enhance Language Awareness and communicative ability.

Recruiting University: University of Vienna (UNIVIE), Austria

Partner University for the Joint Doctorate: Université de Lorraine (UL), Nancy, France

Objectives:

Digitisation has a massive impact on the availability and usability of language data, but its effect on Language Awareness and communicative ability is only gradually being explored. IRP14 investigates how massive exposure to structured language can enhance Language Awareness and communicative ability. Structured language mainly involves: a) corpora of authentic language in the form of data-driven learning (DDL); b) LLMs that underlie GenAI chatbots. Drawing on a usage-based approach, IRP14 will:

- 1) compare the learning potentials/affordances from different forms of ‘authentic’ language exposure;
- 2) analyse current uses of corpora and GenAI for language learning in formal and non-formal contexts, exploring how the strengths of corpora and GenAI can be best combined;
- 3) develop new practices for combining use of corpora and GenAI in developing language skills, testing these empirically.

Expected Results:

- Identification and inventory of existing resources (e.g., corpora, GenAI, platforms or interfaces) for commonly and less commonly taught languages;
- Compilation of an NLP-ready corpus for a Less Commonly Taught Language ;
- Development of resources and recommendations for personalised or personalisable uses for life-long learning.

Secondment:

- Lexical Computing (LC), Prague (Czech Republic), 4 months

Languages: English, German, French, Spanish